The Battle of Neighborhoods - Vegetarian Restaurant in Toronto

Nastaran Dashti

December 12, 2020

1 Introduction

Toronto is one of the most densely populated areas in Canada with an estimated population of over 6 million. Being the land of opportunity, it brings in a variety of people from different ethnic backgrounds to the core city of Canada, Toronto. A new study from HappyCow [1] has also determined which of the world's cities are the most vegan and vegetarian friendly, and Toronto is in the world's top five. The study praises the improvement in the Canadian city's vegan scene, saying it has "come a long way in the past few years and continues to grow."

In this report I will attempt to answer the questions "Where should I open an vegetarian restaurant?" In order to answer this question, this report will try to gather data about Market Places, competition in particular location, aiding places that make people come to restaurants, Population.

This project is aimed towards Entrepreneurs or Business owners who want to open a new vegetarian Restaurant or grow their current business. The analysis will provide vital information that can be used by the target audience

2 Data

The data will provide the list of neighborhoods in Toronto (via Wikipedia), the Geographical location of the neighborhoods (via Geocoder package) and Venue data pertaining to vegetarian restaurants (via Foursquare). The Venue data will help find the best neighborhood to open an vegetarian restaurant.

For the Toronto neighborhood data, a Wikipedia page[2] exists that has all the information we need to explore. It include the postal code, borough and the name of the neighborhoods present in Toronto. We need to scrape the Wikipedia page and wrangle the data and then read it into a pandas dataframe.

In order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood. We will use the Geocoder Python package [3]

All data related to locations and quality of vegetarian restaurants will be obtained By using FourSquare API.

3 Analysis of the data

First we merge data from first and second sources. As it is shown in Fig 1 the features we have are; borough, Neighborhood, Latitude, Longitude.

	Borough	Neighborhood	Latitude	Longitude
0	North York	Parkwoods	43.753259	-79.329656
1	North York	Victoria Village	43.725882	-79.315572
2	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Figure 1: Toronto neighborhood data set.

We can look at the distribution and number of neighborhood for each borough in Figs 2 and 3. We see that North York has highest number of neighborhoods.

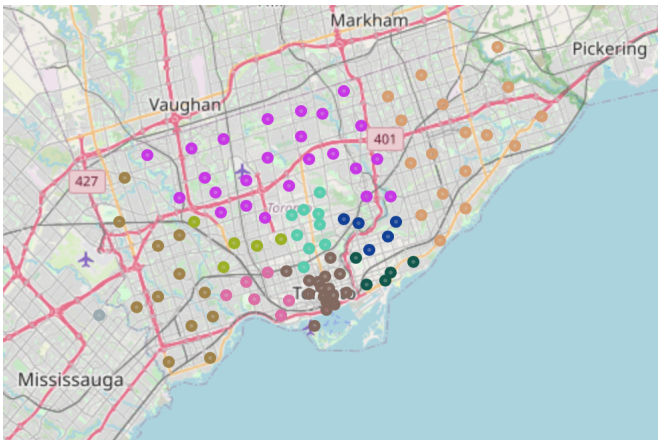

Figure 2: Toronto map with colour-coded each Neighborhood depending on what Borough it was located in.

Next, we used the Foursquare API to get a list of all the Venues in Toronto which included Parks, Schools, Café Shops etc. Getting this data was crucial to analyzing the number of vegetarian restaurants all over Toronto. We then merged the Foursquare Venue data with

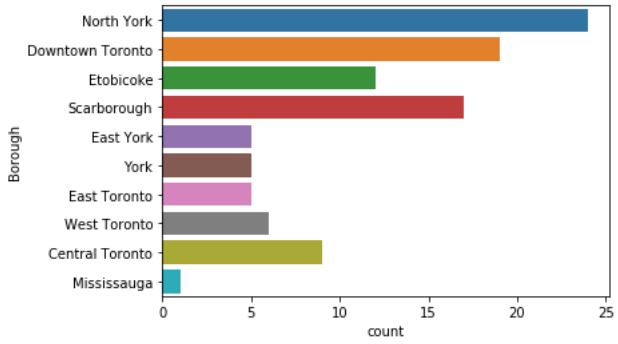

Figure 3: Number of neighborhood for each borough.

the Neighborhood data which then gave us the nearest Venue for each of the Neighborhoods. for illustration purposes, we show venues of the neighborhoods in North York in fig 4.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

Figure 4: North York venues data set.

4 Machine Learning Models

One of the algorithms that can be used here is K-means clustering. K-means is a type of partitioning clustering. That is, it divides the data into k non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm. Objects within a cluster are very similar, and objects across different clusters are very different or dissimilar. We want to cluster the neighborhoods based on the neighborhoods that had similar averages of vegetarian Restaurants in that Neighborhood.

Using FourSquare API we find all venues for each neighborhood. Then, the technique is called One hot encoding, is used to transform Categorical Data to Numerical Data for Machine Learning algorithms. For each of the neighborhood, individual venues were turned into the frequency at how many of those Venues were located in each neighborhood. Finally, we created a new data frame that grouped by Neighborhood names as well as the mean frequency of vegetarian restaurants in that Neighborhood, see Fig **??**.

For K-means clustering, we need to find the optimal number of clusters, which is k. The K-Elbow Visualizer implements the "elbow" method of selecting the optimal number of k. The elbow method runs k-means clustering on the dataset for a range of values for k (from 1-10) and then for each value of k computes an average score for all clusters. If the line chart resembles an arm, then the "elbow" (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In our model the best k is 4 which is shown in Fig **??**.

	Neighborhoods	Vegetarian / Vegan Restaurant
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.000000
5	Berczy Park	0.018182

Figure 5: Data set of the mean frequency of Vegetarian Restaurants in each Neighborhood.

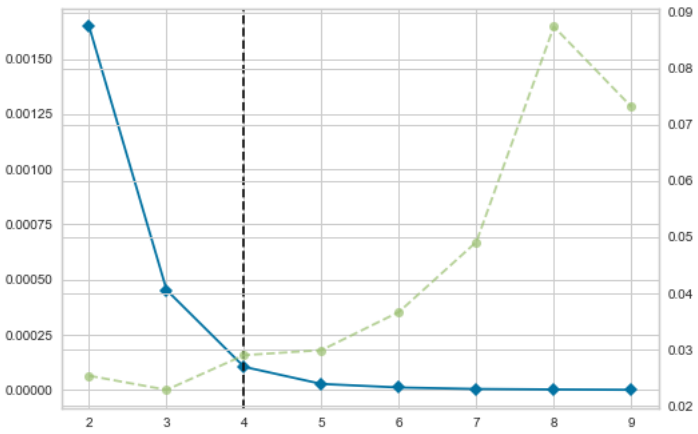

Figure 6: Finding the best k via elbow method.

So then we run our model and group our neighborhoods into four clusters. Each of these clusters was labeled from 0 to 3. Similarity in each clusters are based on mean frequency of restaurant. Fig **??** shows each neighborhood based on the cluster label.

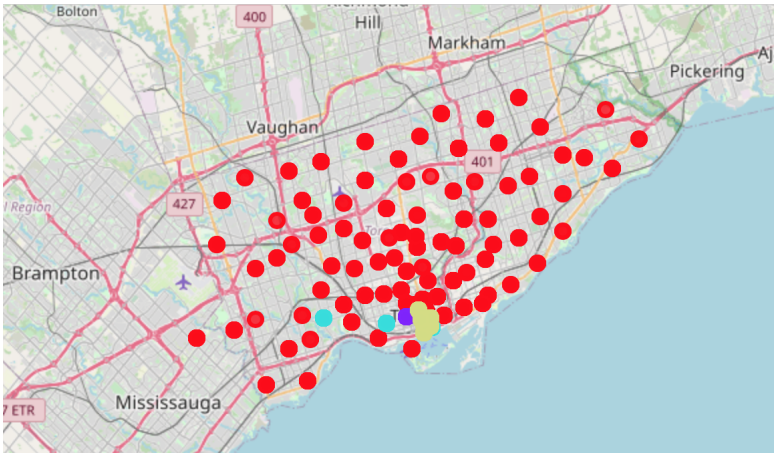

Figure 7: Toronto map with colour-coded each Neighborhood based on the cluster label.

4.1 Results

We can compare the number of Neighborhoods per Cluster Fig **??**. We see that Cluster 1 has the most neighborhoods, 84, while cluster 2 has the least, 1. Cluster 3 has 4 neighborhoods and cluster 4 has only 7.

Then we compared the average Vegetarian Restaurants per cluster Fig **??**. We can see that

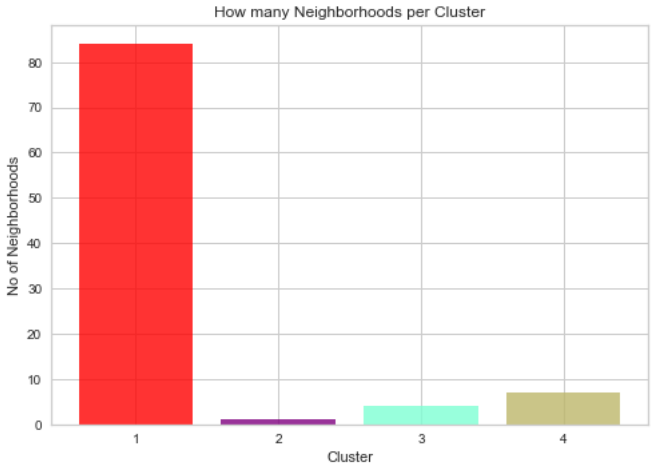

Figure 8: Number of Neighborhoods per Cluster.

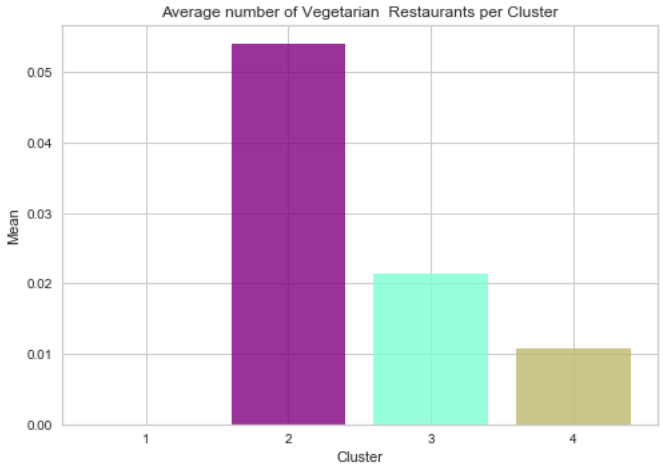

Figure 9: Average Vegetarian Restaurants per cluster.

even though there is 87 neighborhoods in Cluster 1, but it has the least average of Vegetarian Restaurants which is zero. However, in cluster 2 which has only one neighborhood, there is highest number of Vegetarian Restaurants, 0.05.

5 Conclusion and Discussion

Most of the Vegetarian Restaurants are in cluster 2. Even though there is a huge number of Neighborhoods in cluster 1, there is almost no Vegetarian Restaurants.

The optimum place to open a new Vegetarian Restaurant is neighborhoods in cluster 1 which there are many Neighborhoods in the area but no Vegetarian Restaurants, therefore, eliminating any competition.

Since there are not many Vegetarian Restaurant in Toronto and since this city is one of the most vegan and vegetarian friendly city in the world, many of the neighborhoods of the city has the potential of opening Vegetarian Restaurant. We can also go further and look at rating and tips of the restaurant to see how we can compete with those one that already exist.

References

[1] Top Vegan Cities In The World 2019 - HappyCow (2020). URL https://www.happycow.net/vegtopics/travel/top-vegan-friendly-cities. [Online; accessed 7. Dec. 2020].

[2] Contributors to Wikimedia projects. List of postal codes of Canada: M - Wikipedia (2020). URL https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=979555370. [Online; accessed 8. Dec. 2020].

[3] Geocoder: Simple, Consistent — geocoder 1.38.1 documentation (2019). URL https://geocoder.readthedocs.io/index.html. [Online; accessed 8. Dec. 2020].